# Evaluarea modelelor multifactoriale de regresie

# Criterii de evaluare

- Teoretice: compatibilitatea rezultatelor obtinute cu teoria economica (ex.: semnul si marimea parametrilor estimati)
- Statistice:
  - Nivelul de semnificatie al parametrilor
  - Proportia variatiei “explicate” de multimea variabilelor independente
  - Eroarea standard a estimarii
  - Autocorelatia valorilor reziduale
- De previziune

# Metoda regresiei multiple

- Permite analiza relatiei liniare dintre o variabila dependenta si una sau mai multe variabile indepentende
- Obiectiv: explicarea si previziunea variatiei variabilei dependente in functie de covarianta ei cu variabilele independente.

$$\hat{Y} = a + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_i X_i + \ldots + \hat{\beta}_n X_n$$

- Utilizeaza metoda celor mai mici patrate
- Ex.: cererea de bunuri/servicii (dependenta) in functie de factori determinanti (venituri, cifra de afaceri, pret, etc.)

# Semnificatia statistica a parametrilor

- Se dau asigurari ca variatia variabilei dependente nu este datorata intamplari, ci este rezultatul variatiei uneia sau mai multor variabile independente.
- Testul t este utilizat cel mai frecvent.
- Intre nivelul de semnificatie si nivelul de incredere in testul t exista relatia:

Nivelul de semnificatie = (1-nivelul de confidenta)/2

# Eroarea standard

- Eroarea standard a unui parametru estimat arata cu cat poate sa varieze acesta in jurul valorii sale ca urmare a erorii aleatoare.
- Limitele variatiei sunt date de relatia:

$$\beta_j \pm s_{\hat{\beta}_j} \times t_{T,j}$$

# Semnificatia statistica a asocierii dintre variabile

- Testarea semnificatiei legaturii dintre variabila dependenta si variabila/ele independente se bazeaza pe utilizarea testelor statistice (ex.: testul Fisher).
- Se explica in ce masura variatia totala a variabilei dependente este rezultatul variatiei variabilelor independente considerate.
- Valoarea calculata a lui F foloseste formula:

$$F_c = \frac{(\hat{Y} - \bar{Y})/(k-1)}{(\hat{Y} - \bar{Y})^2/(n-k)}$$

# Caracterizarea multilaterala a intensitatii legaturilor

- Matricea coeficientilor de corelatie simpla prezinta intensitatea legaturilor intre toate perechile de variabile
- Matricea coeficientilor de corelatie partiala descrie intensitatea legaturilor intre doua variabile, excluzand efectul celorlalte variabile implicate.
- Coeficientul de corelatie multipla si coeficientul de determinare caracterizeaza proportia variatiei variabilei dependente datorata variatiei setului variabilelor independente ale modelului si proportia variatiei aleatoare (neexplicate).

# Coeficientul de determinare

- Coeficientul de determinare $R^2$ reprezinta raportul dintre variatia explicata si variatia totala, dupa formula:

$$R^2 = 1 - \frac{\sum e^2}{\sum y^2}$$

unde e reprezinta valoarea reziduala și y abaterea variabilei Y de la media sa $\hat{Y}$

- Utilizeaza metoda celor mai mici patrate
- Ex.: cererea de bunuri/servicii (dependenta) in functie de factori determinanti (venituri, cifra de afaceri, pret, etc.)

# Analiza multicoliniaritatii

- Coliniaritatea reprezinta relatia liniara dintre doua variabile independente ale unui model.
- Prezenta sa poate duce la distorsiuni serioase ale parametrilor modelului.
- Sugerata de prezenta erorilor standard mari sau de sensitivitatea exagerata a parametrilor.
- Evidentiata utilizandu-se cele trei teste Farrar si Glauber.

# Primul test Farrar si Glauber

- Se bazeaza pe compararea matricei de corelatie $Z^TZ$ a modelului cu matricea unitate, cu ajutorul testului $\chi^2$

$$\chi_c^2 = -\left[n - 1 - \frac{1}{6}(2(m-1)+5)\right]\ln\det[Z^T Z]$$

- Valoarea teoretica a lui $\chi^2$ se regaseste in tabelele statistice ale repartitiei $\chi^2$, considerandu-se $V=1/2(m-1)(m-2)$ grade de libertate.
- Daca $\chi^2 > \chi^2$, atunci se concluzioneaza ca exista multicoliniaritate la nivelul modelului (regresiei) analizate.

# Al doilea test Farrar si Glauber

- Permite identificarea variabilelor cel mai afectate de coliniaritate
- Se bazeaza pe compararea matricei de corelatie $Z^TZ$ a modelului cu matricea unitate, cu ajutorul testului Fisher.

$$F_c = (r^{ii} - 1)\frac{(n-(m-1))}{m-2}$$

- Valoarea teoretica a lui F se regaseste in tabelele statistice ale repartitiei Fisher, considerandu-se n-m+1 so m-2 grade de libertate.
- Daca $F_c > F_t$, atunci se concluzioneaza ca ipoteza ortogonalitatii intre variabilele independente nu este acceptata.

# Al treilea test Farrar si Glauber

- Permite stabilirea semnificatiei statistice a coeficientilor de corelatie
- Coeficientii de corelatie partiala intre Xi si Xj se determina pe baza formului:

$$r_{ij} = \frac{-r^{ij}}{\sqrt{r^{ii}} - \sqrt{r^{jj}}}$$

- Apoi se calculeaza valoarea testului Student dupa formula:

$$t_{ij} = \frac{r_{ij} \times \sqrt{n-(m-1)}}{\sqrt{(1-r_{ij}^2)}}$$

- Daca $t_{ij} > t_t$, atunci se concluzioneaza ca ipoteza nula este respinsa.

# Analiza erorii medii patratice a valorilor reziduale

- Masura sintetica a acuratetii modelului si o metoda de evidentiere a erorilor de previziune.

$$\frac{1}{T}\sum_{t=1}^{T}(P_t - A_t)^2 = (\overline{P} - \overline{A}) + (S_P - S_A)^2 + 2(1-r)S_P S_A$$

- $(P\text{-}A)^2$ indica tendinta medie a modelului de a supraestima sau subestima valorile reale.
- $(S_P\text{-}S_A)2$ indica sensitivitatea modelului la modificarea valorilor independente.
- $2(1\text{-}r)S_P S_A$ indica marimea erorii datorate lipsei corelatiei perfecte dintre valorile previzionate si cele actuale.

# Analiza autocorelatiei

- Testul Durbin-Watson necesita calculul parametrului d, dupa formula:

$$d = \frac{\sum_{t-2}^{T} (\hat{U}_t - \hat{U}_{t-1})^2}{\sum_{t=1}^{T} \hat{U}_t^2}$$

- Daca $d < d_L$ sau $d > d_T$, atunci este acceptata ipoteza nula ($d_L$ si $d_T$ sunt luate din tabelele asociate testului Durbin-Watson).
- Testul Geary este de natura neparametrica si are ca punct de plecare calculul numarului schimbarilor de semn in seria valorilor reziduale δ.
- Daca $\delta_{min} < \delta < \delta_{max}$ (tabelate), atunci ipoteza nula este acceptata.